

OCR von Inkunabeln

Herausforderungen und Herangehensweisen

Inkunabeln stellen nicht nur die früheste Phase modernen Buchdrucks seit Gutenberg dar (1450-1500), sondern spiegeln als literarische Dokumente auch eine höchst turbulente Umbruchsphase der europäischen Gesellschaft wider, in der alte Gewißheiten in Frage gestellt und neue Horizonte erschlossen wurden. Die elektronische Erschließung dieser typografisch anspruchsvollen, oft in Fortsetzung der Manuskripttradition bewußt künstlerisch gestalteten und teilweise später nicht wieder aufgelegten Werke verspricht deshalb für eine Vielzahl von Geisteswissenschaften einen reichen Ertrag, der sich insbesondere dann ergibt, wenn die reinen Seitenbilder auch durch elektronisch lesbaren und durchsuchbaren Text ergänzt werden.

Bis vor kurzem war es Konsens, dass Inkunabeln wegen der individuellen und mit zahlreichen Abkürzungsglyphen versehenen Typografie einer Behandlung durch OCR nicht zugänglich seien (Rydberg-Cox 1009, <http://digitalhumanities.org/dhq/vol/3/1/000027/000027.html>). Neuere Ansätze einer auf individuelle Typografien trainierbaren, auf neuronalen Netzen aufbauenden OCR rücken die maschinelle Schrifterkennung dieser frühesten Druckerzeugnisse in greifbare Nähe. Der Vortrag stellt Methoden vor, mit denen den spezifischen Herausforderungen begegnet werden kann, zeigt anhand von Beispielen die bisher erreichte Texterkennungsgüte auf und macht Vorschläge für ein koordiniertes Vorgehen zur Hebung des in den Inkunabeldrucken schlummernden Kulturschatzes.

Dr. Uwe Springmann

Centrum für Informations- und Sprachverarbeitung (CIS)

Ludwig-Maximilians-Universität München